

Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of
Business

Lee Kong Chian School of Business

9-2017

Integrating situational judgment tests and assessment centre exercises into personality research: Challenges and further opportunities

Filip LIEVENS

Singapore Management University, filiplievens@smu.edu.sg

DOI: <https://doi.org/10.1002/per.2119>

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the [Human Resources Management Commons](#), and the [Organizational Behavior and Theory Commons](#)

Citation

LIEVENS, Filip. Integrating situational judgment tests and assessment centre exercises into personality research: Challenges and further opportunities. (2017). *European Journal of Personality*. 31, (5), 487-502. Research Collection Lee Kong Chian School Of Business.

Available at: https://ink.library.smu.edu.sg/lkcsb_research/5731

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Integrating situational judgment tests and assessment centre exercises into personality research: Challenges and further opportunities

European Journal of Personality. 2017. Vol. 31, Issue 5, pp. 487-502

<https://doi.org/10.1002/per.2119> [Open peer commentary and author's response, pp. 441-502]

By FILIP LIEVENS

Ghent University, Ghent, Belgium

filip.lievens@ugent.be

Abstract: In this rejoinder, I discuss several broad themes that emerged from the 32 commentaries in response to my target article about the value of integrating Situational Judgment Tests and Assessment Centre exercises for advancing research on the personality–situation interplay. Specifically, I summarize the various challenges put forward (e.g. deficient/contaminated trait measurement, assessment of situation construal, inclusion of personality-driven situation experience) and offer potential conceptual (e.g. use of situational taxonomies and contextualized personality perception) and methodological solutions (e.g. use of open-ended and nonlinear SJTs and multiple speed assessment). Taken together, this fits into a more comprehensive multi-method approach to personality assessment.

When I was preparing the first version of my article as a key-note for the European Conference on Personality in Lausanne, some of my industrial and organizational psychologist colleagues suggested that I was entering a lion's den. After all, I was speaking (as an industrial and organizational psychologist) about personality assessment to an audience of personality researchers. But being open to new experiences ('Every day do something that scares you'), I was looking forward to this opportunity. I also tremendously enjoyed the constructive and enriching interactions after the keynote. When the comments on the written version of my article trickled in, the same positive vibes returned. I am therefore very grateful to all authors who wrote such thoughtful, insightful, and open-minded comments. I agree with many of the ideas expressed in them and also discovered many useful new ones.

The purpose of my article was to set up a constructive dialogue about the viability of incorporating Situational Judgment Tests (SJTs) and Assessment Centre exercises (ACs) in personality research. I concentrated on these two procedures because they have yet to be taken up by personality researchers and because they match well with a focus on personality–situation interplay. In particular, I aimed to show how these assessment approaches permit tackling key issues such as within-person variability across situations, trait-behaviour linkages, and trait expression and perception.

The dialogue got off to an excellent start with the 32 comments. They also provided a first important test of the viability of integrating the two approaches in personality research. Throughout, there was general agreement on many points. First, scholars mentioned that personality psychology has remained too insulated from industrial and organizational psychology and that cross-fertilization between these domains has considerable potential to extend knowledge on the aforementioned topics. Second, need for more diversity in the existing methodological approaches for assessing personality was acknowledged. Third, the emphasis on ecological validity and on behaviour in real-world settings in personnel selection was applauded. Finally, the two assessment procedures were regarded as providing unique opportunities for better understanding the personality triad (persons, situations, and behaviours). At the same time, the

commenters also noted challenges to be met, barriers to be overcome, and additional opportunities for research and collaboration. In this rejoinder, I am not able to address all specific issues mentioned throughout the individual comments. Instead, I focus on several broad themes that emerged from them.

Should One Put Effort into Contextualizing Personality Assessment?

A common element in SJTs and ACs is that they assess traits in more contextualized formats. Selection researchers started adopting these methods (beyond self-reports) because they have potential to lead to better predictions, to be less susceptible to faking good, and to receive more favourable applicant perceptions. Some commenters supported this contextualized approach with additional arguments. For example, Durbin and Hicks applauded the contextualized approach inherent in SJT/ACs because it is both deep (indicative of hypothetical response choices in SJTs and of behaviour in ACs) and narrow (responses relate to particular life situations). In addition, Baumert and Blum made a case that reliance on contextualized items from a variety of contexts is more likely to capture a construct in its full range than self-reports of global traits because the latter require people to average their behavioural tendencies mentally across many situations. Hopwood and Bleidorn mentioned that contextualized measures and contextually relevant traits might be more useful to predict personality changes than applying broad measures multiple times.

Other commenters raised critical voices about the need for contextualization. Möttus argued it is premature to invest in newer more contextualized measurement methods. Instead, he suggested looking for the best possible uses of already existing methods to improve predictive accuracy. Specifically, he pleaded for development and validation of item-based predictive models via machine learning (instead of using a priori composite trait scores). Although this view deviates from how personality has been measured for decades and might open the door to dust-bowl empiricism, I agree that incorporating machine learning in personality measurement is one of the main opportunities in the years to come. In addition, I stress that machine-learning approaches do not preclude using contextualized personality items (alongside generalized ones).

Christiansen and Speer cautioned that increased contextualization in SJTs and ACs might lead to lower validity because the context decreases bandwidth (reflecting a narrower range of demands) and/or the judges are less accurate. Funder argued that the goals of personality psychology and applied domains such as industrial and organizational psychology (personnel selection) are not the same: personality psychology focuses more on understanding than prediction and therefore benefits from using global traits that go beyond specific contexts, whereas the opposite is true for more applied domains. Funder argued therefore that too much contextualization might jeopardize the explanatory value of personality; a point that was also made by Leikas.

I believe Funder's distinction between prediction and explanation should be given full attention. Yet I am hesitant to equate personality and industrial/organizational psychology with the two extremes of his continuum because neither discipline maintains (e.g. social effects of personality in the personality domain) a pure focus on either explanation or prediction. In addition, research demonstrates that contextualizing personality inventories does not necessarily lead to poorer prediction of broad outcomes. On the contrary, meta-analytic research revealed that contextualized personality self-reports predicted overall job performance twice as well than did global non-contextualized self-reports (Schaffer & Postlethwaite, 2012). In any case, I echo the calls of many commenters for more research on contextualized personality measures in various contexts.

SJTs in Personality Research: Challenges and Opportunities

My article put a heavy emphasis on SJTs as new innovative approaches for advancing personality research on within-person variability, trait-behaviour linkages, and personality disorders. Overall, the commenters concurred that the SJT format is a cost-effective (albeit ecologically less valid) alternative to experience-sampling, thereby allowing more control in examining within-person variability and situation-trait contingencies via the standardization of situations. As noted by Corr, an integrative picture emerges by integrating situationally driven intra-individual differences (variability) with inter-individual differences (diversity), which is consistent with Whole Trait Theory (Fleeson & Jayawickreme, 2015). Commenters also saw potential in using situational taxonomies to build SJT scenarios and examining individual differences in situation construal. This might provide insight into people's mental models and how they relate to (mal)adaptive behaviour. Yet the commenters also pointed out several challenges and opportunities for future research.

Trait Measurement in SJTs

Some scholars expressed concerns about SJT scores' construct-related validity. Essentially, these concerns dealt with deficiency and contamination in them (Judge, Hofmans, & Wille; Ziegler & Horstmann). Ziegler and Horstmann referred to potential contamination of SJT scores by crystallized intelligence. In a similar vein, Judge et al. warned that many SJTs are heterogeneous at the item level so that their scores typically correlate with a variety of constructs that are often not intended to be assessed, thereby making the scores less than clear measures of any personality construct.

I agree that construct measurement in SJTs should not be taken for granted. There are at least two aspects that one should note. First, at a conceptual level, it is pivotal to keep the underlying theory behind SJTs in mind. As delineated in the theory of knowledge determinants of SJTs (Lievens & Motowidlo, 2016; Motowidlo & Beier, 2010; Motowidlo, Hooper, & Jackson, 2006), SJT scores reflect procedural knowledge (in the form of implicit trait policies). So, in line with Motowidlo's comment, I stress again that implicit trait policies are not equivalent to personality traits because a person's implicit trait policy represents his or her procedural knowledge about relations between expressions of traits and their effectiveness in situations to achieve some goal. Motowidlo further clarified that the trait (e.g. agreeableness) is one of the antecedents of the implicit trait policy for a given trait.

Second, there is the measurement aspect. Although I agree that construct measurement was an Achilles Heel of many early SJTs, recent research with construct-driven SJTs provides evidence that it is possible to assess specific well-defined traits (i.e. people's procedural knowledge about the effectiveness of these traits; e.g. Bledow & Frese, 2009; Motowidlo et al., 2006; Mussel, Gatzka, & Hewig, 2016). In particular, I refer to Mussel et al. (2016) for an excellent example. They constructed an SJT with acceptable convergent and discriminant validities between self and other reports of analogous personality traits. Key 'ingredients of their recipe' to design such an SJT included reliance on clearly defined personality traits instead of vague dimensions and development of response options to reflect low versus high trait levels (see Lievens, in press for a more general overview of construct-driven SJTs).

Situation Construal and SJTs

Several commenters shared my enthusiasm for using SJTs (and especially implicit trait policies) as novel approaches for assessing situation construal because situation construal is generally recognized to be a proximal cause of behaviour (Funder, 2016). There was agreement that measuring it via SJTs might improve explanation and understanding of people's behaviour, cognitions, and feelings. For instance, according to Wood, Lowman, and Harms, individuals engage every day in a large variety of 'low-fidelity

situation simulations' (e.g. when choosing what clothes to wear) prior to engaging in them, which underscores the relevance of SJTs. Baumert and Blum referred to reactive transactions that can be assessed by varying situational features in SJT scenarios and by examining how people differ in how they perceive the same scenarios (and act upon them when an SJT item's situation occurs in real life). I echo that assessment of situation construal is of pivotal importance in promoting SJT use in the personality domain due to its role in many theories (e.g. CAPS).

That said, some caveats and challenges related to the use of SJTs and situation construal were also mentioned. Motowidlo qualified my suggestion that SJT items can be developed for wide arrays of life domains (e.g. school, health, romantic relationships, and sports). He posited that SJTs' notion of implicit trait policies would be most useful for predicting goal-oriented behavioural patterns that are intrinsically functional or dysfunctional for individuals or society.

Wood et al. argued that implicit trait policies and situation construal as captured via SJTs do not represent psychological situations in their full ranges. They proposed the notion of field representations to capture psychological situations more comprehensively. People's functional field representations of the situation were defined as the mental simulations and expectancies about how situation features covary to determine how to respond to the situation. To assess functional field representations via SJTs, Wood et al. proposed prompting people not only about actions they would choose but also about their evaluations of how their chosen actions would affect specific situation aspects. Similarly, Durbin and Hicks stressed that SJT measurement should be widened by incorporating people's schemas/expectations regarding the likely outcomes of selected responses.

To improve assessing respondents' situation construals further, Rockstuhl made a strong plea for development and use of open-ended SJTs. Such SJTs do not present lists of predetermined response options to individuals. Instead, they ask people for free responses. This can be done in writing (written constructed responses), orally (via think-aloud, see Durbin & Hicks), or via patterned behaviour description interviews (see Heimann & Ingold). Moreover, Rockstuhl posited that situational construal can best be assessed when SJTs manipulate specific situational cues on the basis of situational taxonomies and that such an approach might expand understanding of situation–trait contingencies.

I agree that extensions of the typical SJT format such as open-ended SJTs (especially if they are coupled with manipulations of situational features) are useful to focus on how people construe the standardized scenarios, their memories of similar situations, and their rationales for different responses. At the same time, I note that multiple-choice format does not per se preclude assessing situation construal. It is also possible to measure situation construal via multiple-choice format (e.g. How do you judge the situation? Choose one of the four responses below).

Personality-Driven Situation Experience and SJTs

Several scholars (Lezotte, Condon, & Mroczek; Lilienfeld; Wood, Lowman, & Harms) mentioned the 'contrived' nature of SJTs as a potential limitation. In particular, a common thread running through many commentaries (Baumert & Blum; Judge, Hofmans, & Wille; Lilienfeld; Oh, Kim, & Kim; Rauthmann) was that SJTs typically present people with pre-determined sets of SJT items. Thus, all people receive the same standardized situations in the same sequence, which does not match how actual situations unfold. I acknowledged this limitation inherent in the traditional make-up of SJTs in my article. Thus, I concur that one of the greatest advantages of SJTs, their standardized stimulus and response format, might also be their biggest liability. I also agree that the prototypical SJT mainly gets into personality reactivity (Judge et al.) and suppresses variance in personality-driven situation experience (Rauthmann, Sherman, Nave, et al., 2015) because it constrains individuals' propensities to select situations, shape existing situations, and

create novel situations. As discussed by Judge et al., the standardization imposed by SJTs might also present situations to people that are not representative of those they typically encounter. According to Oh et al., this drawback is especially important for assessing personality traits such as proactive personality (but see the personal initiative SJT developed by Bledow & Frese, 2009).

I propose at least two approaches to deal with this limitation. First, I suggest adopting nonlinear SJTs in which respondents' answers to earlier items influence the subsequent situation items they receive. Baumert and Blum also proposed this. Conceptually, they argued that non-linear/branched/game-like SJTs might get beyond reactive transactions and dig deeper into proactive/manipulative/evocative transactions because they would offer opportunities to enter, avoid, or shape specific situations. As a potential downside, this non-linear and gamified approach challenges adequate construct measurement in SJTs even more. I concur with Handler (2013) that this trade-off between realism and trait measurement will be one of the key challenges in the years to come. Or as Handler aptly put it: 'The use of simulations forces us to choose between raw empiricism that does not provide sound trait-based measurement and highly structured and less fluid simulations, that while measuring important traits, place limitations on realism and complexity. I believe that the future lies in bridging this gap.' (p. viii).

Second, as SJTs are low-fidelity simulations of what people say they will do in different situations (instead of what they actually do), Breil, Geukes, and Back suggested using SJTs and ACs in tandem as a promising means to understand intra-individual variability better in controlled settings. They proposed comparing discrepancies between people's procedural knowledge (assessed via SJT) and their actually expressed behaviour (assessed via AC exercises or webcam SJT).

Other Personality Levels and SJTs

According to Dunlop and Horton, SJTs are based on the narrow assumption that personality primarily consists of trait-based characteristics. They argued that the SJT paradigm should be extended to include other personality aspects. I agree that SJTs focus on personality traits. However, there exist various options to widen the SJT paradigm to include other aspects such as motivations/goals and narratives. Regarding narratives, Dunlop and Horton proposed that SJTs should also identify the stories people associate with the scenarios. This is indeed possible. In fact, in personnel selection, such elaboration has been used to reduce faking good on SJTs (Lievens & Peeters, 2008). Regarding motivations/goals, Dunlop and Horton suggested optimizing SJTs by also identifying the motivations/goals people have for the scenarios. Drawing on the notions of multi-finality (i.e. a behaviour usually has more than one potential consequence) and equi-finality (i.e. different behaviours often have identical consequences), Pretsch and Schmitt made similar suggestions to extend the SJT format by explicitly crossing several traits with several goals in SJT items. Breil et al. proposed designing SJT response options so that they represent different strategies in goal achievement. I am grateful for these suggested extensions because they not only permit obtaining insight into the specific goals that motivate people to choose different trait-related responses but also allow studying rigidity or flexibility in people's trait expressions across goals and situations.

Some commenters (Christiansen & Speer; Heimann & Ingold) went one step further and suggested instruments other than SJTs that might be even better suited to tap into people's goals/motivations and narratives. Christiansen and Speer posited to use biographical questionnaires as alternative personality measures. When these biographical inventories are open-ended (e.g. in the form of essays), I agree that they might further widen the spectrum of personality instruments and tap into personality levels other than traits (i.e. narratives). Yet to obtain insights into the personality–situation interplay, it is crucial that these biographical questionnaires be contextualized to include situational demands that correspond to the demands that affect the behaviours one wants to explain/predict.

Heimann and Ingold made a compelling case for including patterned behaviour description interviews in the personality-research toolbox. In these interviews, people are asked about their trait-related behaviours in situations that they experienced in the past. Unlike SJTs, patterned behaviour description interviews apply an open-ended format wherein participants present situations they encountered and interviewers use prompts to get clear pictures of how participants behaved in them. Patterned behaviour description interviews have several advantages. Apart from self-descriptions of behaviour, the short narrative presented in the interviews provides also information on people's cognitions, emotions, and motivations for choosing specific behaviours and the goals they aim to achieve. People's situation selections also tell something about their personalities. Finally, these interviews allow assessing actual communication behaviour.

Given that such interviews might provide information on the three levels of personality mentioned by Dunlop and Horton, I agree that these interviews further diversify personality assessment methods. That said, I also note that the benefits of patterned behaviour description interviews come with a price of lower stimulus presentation (each interviewee presents different situations and interviewers vary in prompt level and content) and response scoring consistency (interviewers rate people's answers). I am also less enthusiastic about the viability of patterned behaviour description interviews for studying within-person variability because such interviews typically tap fewer situations than do SJT items.

Personality disorders and SJTs

Generally, my suggestions to use SJTs for advancing measurement and understanding of personality disorders were well received. Several commenters (De Fruyt, De Clercq, Verbeke, & Vergauwe; Durbin & Hicks; Wood et al.; Wright) concurred with the various advantages that I mentioned when using SJTs in this domain as additions or even alternatives to structured clinical interviews, self-report questionnaires, and ambulatory assessment approaches. SJTs' advantages include elicitation of individual differences in how patients interpret and respond to sets of standardized scenarios, use of situational taxonomies (e.g. interpersonal circumplex and DIAMONDS) for building these scenarios, ease of administration, and automated scoring. I echo the commenters that SJTs can reveal relevant diagnostic information (and eventually assess personality disorders) by detecting someone's (i) endorsement of overtly problematic response options, (ii) situation–trait contingencies that mismatch those of ‘normal’ people, and (iii) discrepancies between responses to ‘what should you do’ and ‘what would you do’ instructions. Future research should compare the convergence among these approaches in detecting personality dysfunctions.

It was encouraging that De Fruyt et al. have already developed an SJT for assessing borderline personality pathology. This confirms the feasibility of the SJT approach in this area. Durbin & Hicks suggested adding to SJTs think-aloud procedures (for digging into situation construal) and/or behavioural observation (for examining whether people with more maladaptive personality structure have greater discrepancies between knowledge of effective trait-related behaviour and actual expressions of it). Relatedly, Lilienfeld posited that the standardized nature of AC exercises might avoid the lower ecological validity of SJTs and test even better the notion that many personality disorders are inherently interpersonal. Lilienfeld made an especially compelling case for using multiple AC exercises to ensure that the positive impressions that for example narcissists tend to create in the short run (Back, Schmukle, & Egloff, 2010) do not deceive observers.

Finally, at a conceptual level, Wright made a point that I should have highlighted more in my article. He argued that the viability of SJTs in this area is contingent upon a workable definition of personality pathology. That is, SJTs are especially useful if one conceives personality dysfunctions as mismatches between people's trait expressions interact with situational demands (instead of extreme trait levels per

se). I concur with Wright that such a definition determines how personality disorders manifest themselves SJT responses (e.g. rigidity across different situations, instability across similar situations, or situation–trait contingencies that deviate from normative/adaptive patterns).

Importance of relying on an underlying theory of SJT development is also evident in Jayawickreme's proposal for using SJTs to identify individuals who are more likely to experience positive changes after traumatic experiences. For SJTs to predict this, theories about post-traumatic growth should guide their development. Theorizing in this domain informs which adverse situations to sample, which response options to include, and how to score participants' responses (e.g. people that systematically choose options in which they seek social support receive higher scores). When experience-sampling studies serve as criterion measures, an additional benefit of such theory-driven SJTs is that they allow testing specific theoretical assumptions underlying post-traumatic growth.

Quantification of within-person variability in SJTs

Although my idea for using SJTs for capturing within-person variability was favourably received, questions were raised about how to quantify and model this variability (Breil et al.; Corr; Lang, Tackett, & Zettler). Along these lines, Lang et al. provided excellent analytical recommendations. They made a case that using the traditional standard deviation (*SD*) as a within-person variability index in SJTs might confound various sources of variance: the *SD* might not only reflect degree of variability with which a person shows a tendency to choose responses related to a given trait across situations; it might also be affected by the social presses of the situations (items), the respondee's trait levels, and individual differences in careless responding.⁷

To avoid these confounds Lang et al. argued convincingly in favour of using Item Response Theory (i.e. tree models) for modelling within-person variability. The key advantage of this IRT-based approach is that it splits information from SJT responses into (i) persons' mean-trait levels/latent traits, (ii) persons' tendencies to express the trait variably, (iii) items' mean-level/latent difficulty, and (iv) items' tendencies to elicit rating variability. As a key implication, this IRT-based approach (as opposed to the classic *SD* approach) permits conceptual clarity because it disentangles people's intra-individual variability from other sources of variance. I thus fully concur that advanced psychometric models represent important and complementary building blocks for research on within-person variability via SJTs. Therefore, they should also be integrated in the Breil et al. framework that distinguishes between within-context variability and cross-context variability.

ACs in personality research: Challenges and opportunities

My article made a case that AC exercises might advance research on personality expression, personality perception, and their transactions. Generally, commenters agreed with the areas of research that I identified but also raised some concerns and suggested additional research ideas, which I elaborate on below.

ACs as personality measures

Similar to SJTs, several authors (Johnson; Judge et al.; Leikas; Pretsch & Schmitt) questioned whether AC exercises assess personality. According to Johnson and Leikas, for example, the behaviour displayed by candidates in AC exercises is not reflective of their standings on personality traits. Instead, the behaviours shown reflect their skills in displaying the behaviour that they consider important for the situations (whether specifically what the AC judges are looking for, or what would be appropriate to the situation and they would do in real life). Judge et al. suggested that the behavioural reactions evoked in

these assessments often reflect constructs that do not belong to the personality domain, thereby creating construct contamination.

Relatedly, various authors (Heimann & Ingold; Judge et al.; Mussel, et al.) also raised questions about the construct-related validity of AC ratings. They pointed to the large research base showing that ratings from given AC dimensions do not correlate with ratings of the same dimensions across exercises (e.g. Bowler & Woehr, 2006; Lievens & Conway, 2001). I do not see this robust finding as a limitation. The limited convergence of people's ratings on the same dimension across different ACs is consistent with research on levels of cross-situational consistency across situations when these situations have differing demands (Christiansen, Hoffman, Lievens, & Speer, 2013). In fact, this finding shows that ACs are ideally suited for examining how the interplay of personality and situation shapes people's behaviour.

Trait Activation and ACs

In my article, I discussed how Trait Activation Theory has been fruitfully used to identify which AC exercise features trigger candidate behaviour. Tett offered additional arguments for extending Trait Activation Theory to personality research outside the workplace. As the key principles of trait activation theory outside the workplace have so far been largely untested, he suggested identifying the situational cues (at the task, interpersonal, and contextual levels) that evoke trait-related behaviour in a wide set of life situations (e.g. expression of road rage, prosocial behaviour). Consistent with Rauthman, I also suggest relying on recently developed situational taxonomies for structuring the variety of situational cues. Such taxonomies provide ample opportunities to vary situation cues to evoke differences in trait-related behaviours (Breil et al.). Rauthmann's situation strength matrix is also an excellent example of how personality researchers might vary situations to detect individual differences.

Some commenters, however, tempered enthusiasm for the trait activation logic in AC-like tasks. For instance, Pretsch and Schmitt pointed to potential observability differences among traits (aka hierarchy of trait visibility), thereby questioning whether several simultaneously activated traits are equally visible to observers. They also noted that, especially in group discussions, individuals' trait expressions might depend on other people's trait expressions. I acknowledge this caveat. ACs are diverse and range from single-participant exercises, to one-on-one exercises, to group discussions. The constructs targeted and assessed vary across these formats. When traits are less observable in given exercises, recent research has demonstrated that planting specific stimuli to evoke desired trait behaviours in the exercises (e.g. via role-players or by changing contextual features) and familiarizing assessors with them increases assessor detection and utilization of these behaviours in scoring (Lievens et al., 2015). Breil et al. concurred that role-players and other interaction partners might produce the needed varying and adaptive situational cues (in addition to the fixed ones that are already part of initially presented situations).

In Situ, Ex Situ, and Juxta Situm Assessment

In my article, I did not elaborate on whether raters are present when candidates participate in ACs and/or whether afterwards they watch videotaped candidate performances. Whatever option is chosen in personnel selection, personality researchers could use any combination of various rating approaches. Along these lines, Rauthmann made useful distinctions among three (not mutually exclusive) options: in situ, ex situ, and juxta situm raters. In situ rating refers to participants rating their own performances, whereas ex situ raters are a posteriori raters/coders who evaluate videos of the participants. In current AC practice, combinations of these ratings are common. In online ACs, ex situ raters are often used. On-site assessors who are physically present when the participants participate in the ACs can be regarded as juxta situm raters. In some cases, fellow AC participants and role players might also serve as AC raters.

In prior personality research (e.g. Rauthmann, Sherman, Nave, et al., 2015), *juxta situm* raters were often not available. However, when ACs are used, it is possible (and often recommended) to have combinations of *in situ* raters, *ex situ* raters, and *juxta situm* raters. This triangulation of raters is of pivotal conceptual importance: it enables researchers to disentangle variance due to objective situations, situation construal, and people's traits. In turn, partitioning the variance according to these different factors allows more comprehensive tests of situation construal models (e.g. Funder, 2016).

Personality-Driven Situation Experience and ACs

Contrary to SJTs, AC exercises give people discretion to change and shape situations during the exercises. Unfortunately, current AC practices do not fully capitalize on this advantage because assessors typically rate people's behaviour but not their strategies to shape the situations. Therefore, I echo Rauthmann that it is vital to evaluate the extents to which people change, evoke, or create new situations in AC settings (see also Rauthmann & Sherman, 2016c).

In particular, as discussed in some comments (Gibbons & Rupp; Leikas; Rauthmann), the interpersonal and dynamic features of ACs provide unique opportunities for examining how people shape interpersonal situations and create new ones. Therefore, I concur that ACs (e.g. group discussions and role plays) are ideally suited for testing principles underlying interpersonal theory by systematically varying situational characteristics such as the numbers, status, sexes, intentions, personalities, and so forth of the other participants involved. Thus, the social and dynamic nature of ACs should be viewed as real opportunities. Compared to other instruments (e.g. SJTs), the ecological validity for making inferences related to interpersonal competencies is substantial in ACs, although some losses in internal validity should be acknowledged (Judge et al.; Pretsch & Schmitt).

Multiple Speed Assessments

Some commenters (Lezotte et al.; Lilienfeld; Wood et al.) noted that behavioural variability unfolds over time (over days and weeks) and wondered whether and how this temporal dimension can be included in assessment procedures other than experience-sampling. In my article, I gave the example of how a multiple speed assessment procedure (18 short 3-minute role plays) was set up to assess people's interpersonal competencies (Herde & Lievens, 2016). Per role play, there was one role player who also served as assessor. Such multiple speed assessment enables assessing how people vary their behaviour. Variation can then be conceptualized as behavioural flexibility across the various exercises and/or as 'learning' throughout the session (Baard, Rench, & Kozlowski, 2014).

De Fruyt et al. elaborated on this multiple speed assessment idea. They pointed out how large sets of brief AC-like exercises might be fruitfully used to evaluate personality change after specific interventions (e.g. coaching on expressing specific traits). As an advantage, in this assessment procedure, personality change can be examined independently from the people who received the interventions by making each exercise's assessor a 'blank slate' (i.e. unfamiliar with the participants and blind to the coached traits).

Finally, Fleeson and Hamza built on the multiple speed assessment idea and on the notion of successive winnowing to propose a 'Successive Situation-Based Selection' approach. This intriguing approach permits personality researchers to capitalize on the advantages of behavioural assessment while at the same time reducing some of the typical costs. That is, Fleeson and Hamza suggested gradually reducing initially undifferentiated samples of participants to more extreme (high and low) groups via successive selections on the basis of performance in various ACs. Next, these extreme groups allow researchers examining dimensions of interest (e.g. integrity and dark triad) with more power and less assessment time. Although this procedure does not shed light on variability across time, I agree that it represents

another example of how the AC methodology can be flexibly used. I also anticipate that in the future, social-sensing techniques (Schmidt Mast, Gatica-Perez, Frauendorfer, Nguyen, & Choudhury, 2015) and virtual (remote/online/gamified) ACs might even replace role players/assessors to save costs (see also Lezotte et al.).

Contextualized Personality Perception

In my article, I laid out the many parallels between AC research and the zero-acquaintance literature. I argued in favour of better cross-fertilization between these two literatures that have evolved almost independently. I discussed among other things how ACs and novel developments related to good judgement (dispositional reasoning and transactions between good judges and good targets/good information) might be especially beneficial for personality research.

Generally, commenters (e.g. De Kock; Funder; Gibbons & Rupp; Letzring & Colman; Mignault & Human) reacted enthusiastically to the prospect of embedding the systematic AC into behavioural observation research in the personality domain. Letzring and Colman discussed that such research is needed to test whether results obtained in zero-acquaintance research still stand in real-world situations where the stakes and motivations of both target and judge are higher. According to Corr, the combined attention to personality expression and perception underscores the notion that traits are socially contextualized. Thus, in addition to people's personality expression across situations, how these expressions are judged in ACs provides unique information about targets.

The commenters also built on my suggestions and offered other opportunities for joint research. The most striking common thread was that they suggested adopting more contextualized approaches to personality perception. For instance, De Kock suggested moving from examination of what makes people good judges of others to what makes people good judges of people *in their situations*. He thus argued that contextualized approaches to personality assessment deal not only with focal persons but also with how people judge persons in situations and in person–situation transactions. Only when this part of the assessment process is taken into account will deeper understanding of personality perception be obtained. To this end, De Kock called for developing a measure of situational reasoning (besides the already existing one on dispositional reasoning) and provided various ways researchers might accomplish this. This situational reasoning measure might extend the current contextualization measure that is included in dispositional reasoning (De Kock, Lievens, & Born, 2015).

Gibbons and Rupp also stressed the need for contextualized approaches to personality perception among assessors and in assessor training programs. I agree that current training programs are based on traditional trait psychology and seldom take person–situation transactions into account. Assessors should thus be taught to be more aware of contexts in which behaviours occur. Only then they will be more appreciative of people's *if ... then ...* signatures.

Apart from the advantages mentioned by Gibbons and Rupp, a contextualized approach to personality perception could also inform research on other reports (e.g. from coworkers, family members, and friends). Some scholars (Christiansen & Speer; Oh et al.) noted that such other reports deserved more attention in my article. Although other-reports of target people's personalities, as measures of reputation, have incremental validity over self-reports in predicting important outcomes, this aspect is typically not evaluated as part of the interplay between personality and situation. Interestingly, Christiansen and Speer discussed recent research (Kluemper, McLarty, & Bing, 2017) that attested to the validity benefits of using observer ratings from relevant contexts. We need more studies with other reports that take contextualization into account.

To examine accuracy of such contextualized personality judgements, Letzring and Colman proposed an innovative approach. Contrary to the common practice of measuring person perception accuracy as agreement between target people's standings on given traits and judges' assessments of them, they suggested asking judges to complete SJT items, which ask which course of action target people would choose to respond to the situations. Target people's/acquaintances' responses to the SJT items serve as criteria to determine how accurately judges perceive targets' identities/reputations. I welcome this new approach because (i) number of correctly identified SJT responses represents an objective accuracy benchmark, (ii) it pays more attention to the situations in which personality judgments are made, and (iii) it increases the range of methodologies used in accuracy research.

Finally, Mignault and Human made some suggestions for how research evidence on the good target and expressive accuracy might improve judgement accuracy in personnel selection. I agree that this is largely uncharted territory for selection researchers, who have typically focused on good judge factors. Mignault and Human discussed that it is easier for assessors to rate candidates who can better showcase their qualities. Hence, one recommendation was that assessors should create warm and friendly atmospheres. To improve accuracy Mignault and Human also suggested fostering individuals' self-presentations because research shows it improves people's clarity of expression and judges' cue detection. This runs somewhat counter to common assumptions of selection researchers that self-presentation is often equated with artificial impression management and socially desirable responding. Future studies should pit these competing rationales against each other.

SJTs and ACs in Multi-Method Approaches to Personality Assessment

As many commenters noted (e.g. Connelly & McAbee; Corr; Heimann & Ingold), integrating SJTs and ACs into personality research fits into comprehensive multi-method approaches to personality assessment that capitalize on advantages of all methods included (such as traditionally reliable and cost-effective self-reports⁸).

I endorse the multi-method approach to personality assessment. Yet applying it means that several conceptual, methodological, and statistical questions must be addressed. For starters, it begs the questions what one should do with the traditional measures and, for instance, how one should integrate SJTs and ACs with these measures. I echo Connelly and McAbee's useful distinction between using new measures such as SJTs or ACs as (i) replacements (i.e. using new measures instead of old measures), (ii) supplements (e.g. summing scores across measures to produce more precise and representative multi-method measures), or (iii) complements (i.e. using new measures alongside old measures).

I support the complementary approach because there should not be competition between old and new methods and all measures validly tap varying aspects of similar constructs (see also Durbin & Hicks; Hopwood & Bleidorn). Thus, new assessment methods provide information that partially overlaps with that of traditional ones but also allow assessing previously untapped and unique elements of personality. To this end, Connelly and McAbee outlined a useful structural equation modelling approach to tease out variance shared among measures from variance unique to individual measures.

Conceptually, the complementary perspective is consistent with a broad personality conceptualization that does not equate personality only with traits but also with goals/motivations and narratives (Dunlop & Horton). The complementary perspective is also consistent with a lens model that distinguishes between personality measurements related to (i) behavioural observation, (ii) self-concept, and (iii) interpersonal perception (Back & Egloff, 2009). It is therefore worthwhile to update Back and Egloff's model with the new measures discussed in my article and in the comments (SJTs, ACs, but also patterned behaviour description interviews, biographical questionnaires). Figure 1 presents this updated model.

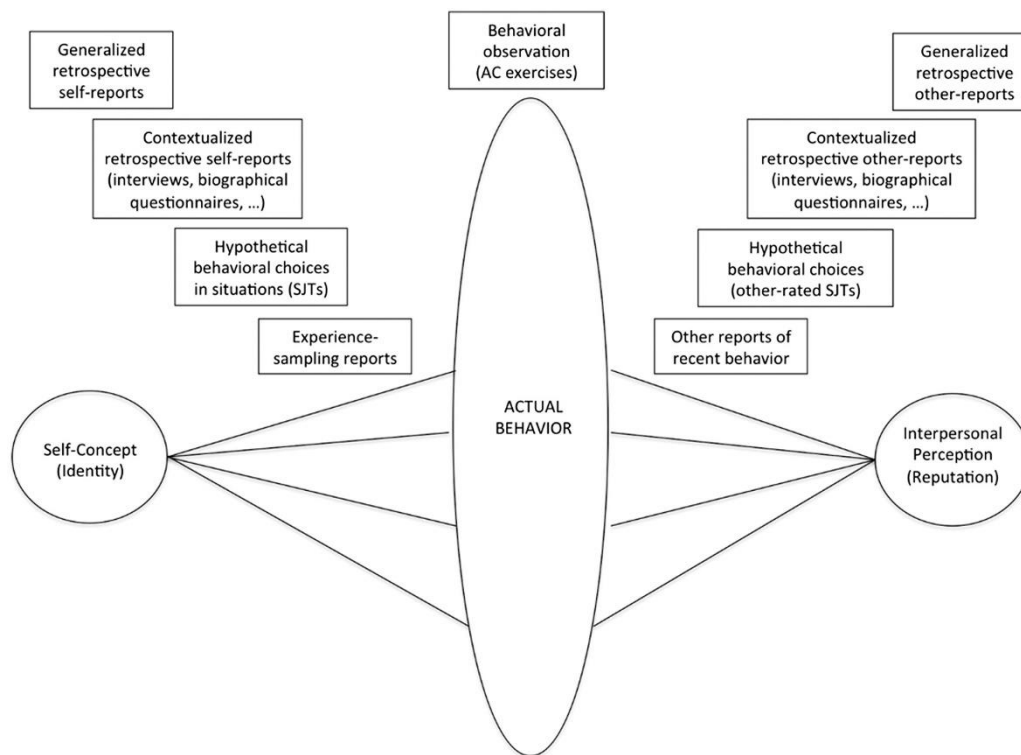


Figure 1

A lens model representation of multiple methods to assess personality (see also Back & Egloff, 2009).

Epilogue

In my article, I showcased two personnel selection procedures (SJTs and ACs) that have largely escaped personality researchers' attention. The subsequent 32 commentaries as well as this rejoinder attest to the start of a constructive dialogue between selection and personality researchers that should enrich the theoretical and empirical approaches of both areas. This fits well into the broad idea that psychology researchers should leave their silos and cross the borders of their own respective disciplines. Once again, I would like to thank all the commenters. Key initial steps have been taken to formulate and refine a joint research agenda. I am therefore confident that both SJTs and ACs will inspire personality researchers in the future and will find their ways into personality research programs.

Notes :

7 Compare this to the *SD* in experience-sampling studies that use the same scales each day. As these scales have larger numbers of items/observations, ratings are less likely affected by situation/item press confounds.

8 I did not intend my article to dismiss self-reports. I simply listed reasons that selection researchers started searching for methods other than self-reports.

References

- Back, M. D., & Egloff, B. (2009). Yes we can! A plea for direct behavioral observation in personality research. *European Journal of Personality*, 23, 403–405.
- Back, M. D., Schmukle, S. C. & Egloff, B. (2010). Why are narcissists so charming at first sight? *Journal of Personality and Social Psychology*, 98, 132–145.
- Baard, S. K., Rench, T. A., & Kozlowski, S. W. J. (2014). Performance adaptation: A theoretical integration and review. *Journal of Management*, 40, 48–99.
- Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative: Towards understanding construct based situational judgment tests. *Personnel Psychology*, 62, 229–258.
- Bowler, M. C., & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology*, 91, 1114–1124.
- Christiansen, N., Hoffman, B. J., Lievens, F., & Speer, A. (2013). Assessment centers and the measurement of personality. In N. Christiansen & R. Tett (Eds.) *Handbook of personality at work*. (477–497). New York (NY): Routledge.
- De Kock, F., Lievens, F., & Born, M. Ph. (2015). An in-depth look at dispositional reasoning and interviewer accuracy. *Human Performance*, 28, 199–221.
- Funder, D. C. (2016). Taking situations seriously: The situation construal model and the riverside situational Q-sort. *Current Directions in Psychological Science*, 25, 203–208.
- Handler, C. (2013). Foreword. In M. S. Fetzer & K.A. Tuzinski (Eds.), *Simulations for personnel selection* (pp. v–ix). New York: Springer.
- Herde, C. N., & Lievens, F. (2016). The added value of speed assessment for assessing leadership: In. P. V. Ingold, & B. J. Hoffman (Chair), AC, Quo Vadis? Novel assessment center design and validation strategies. Symposium conducted at the Annual Conference of the Society for Industrial and Organizational Psychology, Anaheim, CA, USA.
- Kluemper, D. H., McLarty, B. D., & Bing, M. N. (2017). The validity of friend, family, and coworker ratings of personality. Paper presented at the 32nd annual meeting of the Society for Industrial and Organizational Psychology Conference held in Orlando, FL.
- Lievens, F. (in press). Construct-driven SJTs: Towards an agenda for future research. *International Journal of Testing*.
- Lievens, F., & Conway, J. M. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology*, 86, 1202–1222.
- Lievens, F. & Peeters, H. (2008). Impact of elaboration on responding to situational judgment test items. *International Journal of Selection and Assessment*, 16, 345–355.

- Lievens, F., Schollaert, E., & Keen, G. (2015). The interplay of elicitation and evaluation of trait-expressive behavior: Evidence in assessment center exercises. *Journal of Applied Psychology*, 100, 1169–1188.
- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, 91, 749–761.
- Mussel, P., Gatzka, T., & Hewig, J. (2016). Situational judgment tests as an alternative measure for personality assessment. *European Journal of Psychological Assessment*, <https://doi-org.libproxy.smu.edu.sg/10.1027/1015-5759/a000346>
- Rauthmann, J. F., & Sherman, R. A. (2016c). Situation change: Stability and change of situation variables between and within persons. *Frontiers in Psychology*, 6, 1938. <https://doi.org/10.3389/fpsyg.2015.01938>
- Rauthmann, J. F., Sherman, R. A., & Funder, D. C. (2015). Principles of situation research: Towards a better understanding of psychological situations. *European Journal of Personality*, 29, 363–381.
- Schmid Mast, M., Gatica-Perez, D., Frauendorfer, D., Nguyen, L., & Choudhury, T. (2015). Social sensing for psychology: Automated interpersonal behavior assessment. *Current Directions in Psychological Science*, 24, 87–92.